

Correlation and Regression

Experiment-2



June 18, 2021

Bimal parajuli

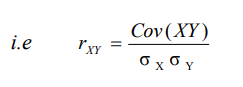
20BDS0405

**Correlation Definition:-**

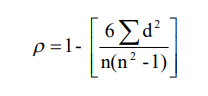
Correlation refers to the relationship between two or more variables. Simple correlation studies the relationship between two variables. Correlation analysis attempts to determine the degree of relationship between variables.

Measures of Correlation:

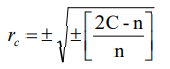
* Scatter Diagram
* Karl Pearson’s Coefficient of Correlation

It is defined as the ratio of covariance between x and y say Cov (X, Y) to the product of the standard deviations of X and Y, say σ (X) and σ (Y) 

* SPEARMAN'S RANK CORRELATION COEFFICIENT

Suppose we associate the ranks to individuals or items in two series based on order of merit, the Spearman's Rank correlation coefficient ρ is given by

* KENDALL'S COEFFICIENT OF CONCURRENT DEVIATIONS

The Kendall's coefficient of concurrent deviations is denoted by rc and defined as 

Where, C = Number of concurrent deviations or position signs of (DX, DY); n = Number of pairs of deviations

**Regression:**

## DEFINITION

Regression analysis is a statistical method of determining the mathematical functional relationship connecting independent variable(s) and a dependent variable.

Its types are:

* Simple linear Regression

In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete and nature of relationship is linear. This relationship can be expressed using a straight line equation (linear regression) that best approximates all the individual data points.

The general form of the simple linear regression equation is Y = a + bX + e, where ‘X’ is independent variable, ‘Y’ is dependent variable, a’ is intercept, ‘b’ is slope of the line and ‘e’ is error term.

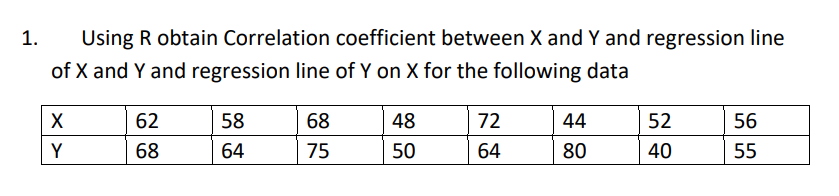
* Multiple linear Regression

Multiple linear regression uses two or more independent variables to estimate the value(s) of the response variable (Y).

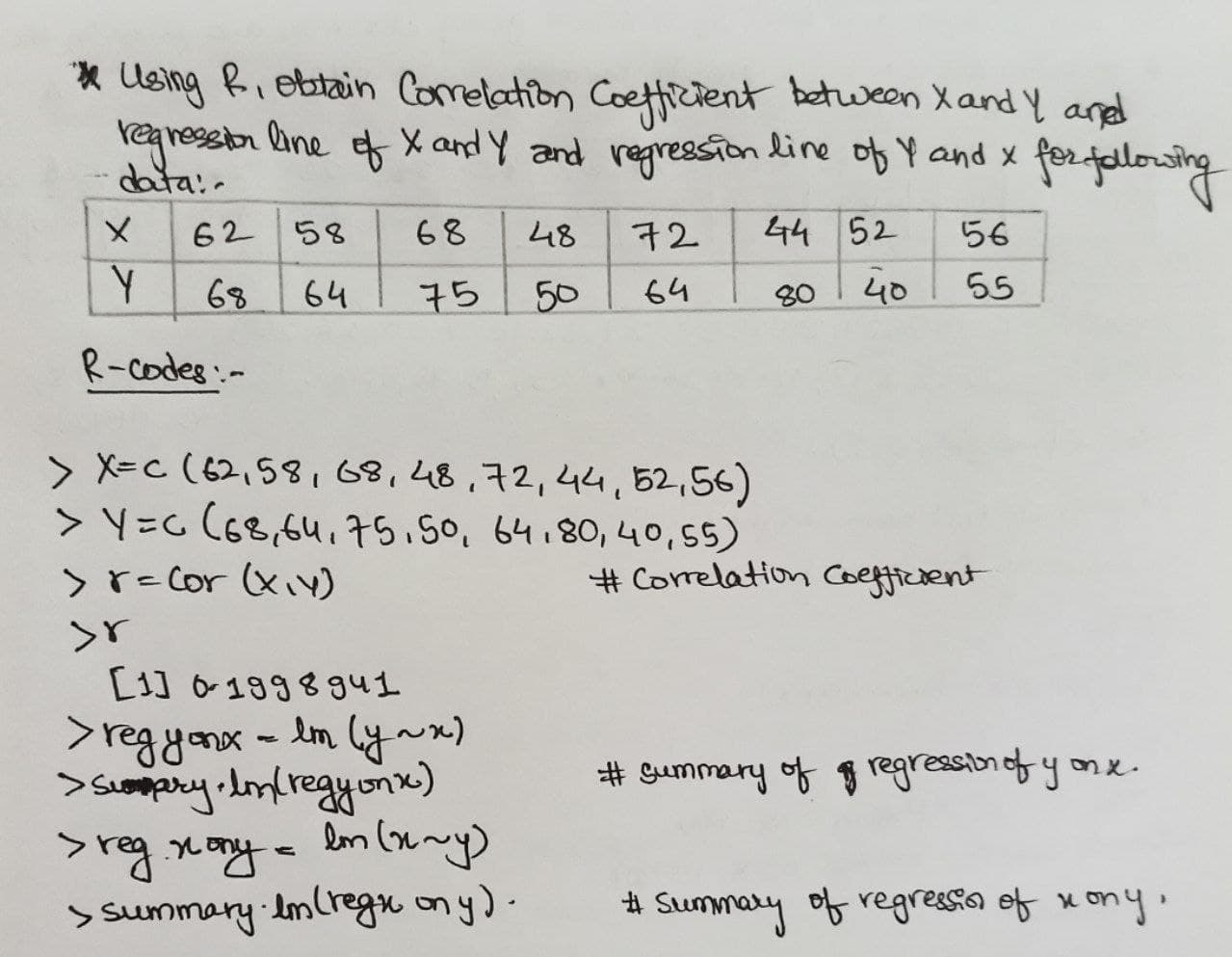
The general form of the multiple linear regression equation is Y = a + b1X1 + b2X2 + b3X3 + ... + btXt + e

* Non Linear Regression

Problem 1:



|  |
| --- |
| **R- Code:** |
|  |



Problem 2:

Calculate the Coefficient of correlation of x and y from the given data:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| X | 23 | 27 | 28 | 28 | 29 | 30 | 31 | 33 | 35 | 36 |
| y | 18 | 20 | 22 | 27 | 21 | 29 | 27 | 29 | 28 | 29 |

|  |
| --- |
| **R- Code:** |

|  |
| --- |
|  |

Problem 3:

Twelve recruits were subjected to selection test to ascertain their sustainability for a certain course of training. At the end of training, they were given a proficiency test. The marks scored by the recruits are recorded below:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Recruit | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Selection Test Score | 44 | 49 | 52 | 54 | 47 | 76 | 65 | 60 | 63 | 58 | 50 | 67 |
| Proficiency test Score | 48 | 55 | 45 | 60 | 43 | 80 | 58 | 50 | 77 | 46 | 47 | 65 |

|  |
| --- |
| **R- Code:** |

|  |
| --- |
|  |

Problem 4:

**The body weight and BMI of 12 school going children are given in the following table. Fit a simple regression model of BMI on weight and examine the results.**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Weight | 15 | 26 | 27 | 25 | 25.5 | 27 | 32 | 18 | 22 | 20 | 26 | 24 |
| BMI | 13.35 | 16.12 | 16.74 | 16.00 | 13.59 | 15.73 | 15.65 | 13.85 | 16.07 | 12.8 | 13.65 | 14.42 |

|  |
| --- |
| **R- Code:** |

|  |
| --- |
|  |